**Biotechnology Letters**

**Characterization of a glucose tolerant beta glucosidase from *Aspergillus unguis* with high potential as a blend-in for biomass hydrolyzing enzyme cocktails**

**Prajeesh Kooloth Valappil.1, 2, Kuni Parambil Rajasree. 1,4, Amith Abraham 1, 3, Meera Christopher 1, 2, \*Rajeev K. Sukumaran 1**

**1***Biofuels and Biorefineries Section, Microbial Processes and Technology Division, CSIR-National Institute for Interdisciplinary Science and Technology, Industrial Estate P.O., Trivandrum 695019, India*

**2** *Ghaziabad, India*

**3***Department of Chemical Engineering, Hanyang University, Seoul, Republic of Korea*

**4** *Present Address: Department of Microbiology and Cell Biology, Indian Institute of Science*

*Bangalore- 560012, India*

**\*** Corresponding Author

Rajeev K Sukumaran, Ph. D

Microbial Processes and Technology Division

CSIR-National Institute for Interdisciplinary Science and Technology,

Thiruvananthapuram- 695 019, India

Tel 91-471-2515368

Fax 91-471-2491712

Email: rajeevs@niist.res.in

**Table S1:** Subcellular localization of putative beta-glucosidases from *A.unguis* identified through Signal P analyses

|  |  |  |
| --- | --- | --- |
| **Protein**  **Id. No.** | **Whether secretory or intracellular** | **Position of Signal Peptide Cleavage Site** |
| G493 | Secretory | 17, 18 |
| G2101 | - | - |
| G2942 | - | - |
| G2944 | Secretory | 19, 20 |
| G3151 | Secretory | 19, 20 |
| G3964 | - | - |
| G5435 | - | - |
| G6460 | Secretory | 23, 24 |
| G7342 | - | - |
| G8848 | - | - |
| G9078 | Secretory | 17, 18 |
| G9081 | Secretory | 17, 18 |
| G10146 | - | - |
| G11844 | - | - |

**Table S2:** Physicochemical properties of secretory BGLs of *A. unguis* determined through *in-silico* analyses

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Protein  ID | No. of amino acids | Molecular Weight (kDa) | Aliphatic Index | Grand average of hydropathicity (GRAVY) | pI |
| G493 | 773 | 84.7 | 78.31 | -0.190 | 5.43 |
| G2944 | 1244 | 135.18 | 72.85 | -0.463 | 5.47 |
| G3151 | 850 | 92.09 | 75.08 | -0.344 | 4.94 |
| G6460 | 823 | 89.57 | 74.31 | -0.346 | 4.89 |
| G9078 | 801 | 84.87 | 72.36 | -0.256 | 4.92 |
| G9081 | 768 | 82.81 | 82.46 | -0.228 | 5.11 |

**Table S3:** Homology modeling of different secretory BGLs from *A. unguis*: Templates and Validation of model quality

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sl. No | Protein ID | SMTL ID | GMQE | QMEAN | Sequence  Identity |
| 1 | G493 | [4iib.1.A](https://swissmodel.expasy.org/templates/4iib.1) | 0.58 | -3.58 | 42.84% |
| 2 | G2944 | [5fjj.1.A](https://swissmodel.expasy.org/templates/5fjj.1) | 0.51 | -2.27 | 47.58% |
| 3 | G3151 | [5fjj.1.A](https://swissmodel.expasy.org/templates/5fjj.1) | 0.91 | -0.41 | 76.41% |
| 4 | G6460 | [5fjj.1.A](https://swissmodel.expasy.org/templates/5fjj.1) | 0.72 | -3.05 | 45.54% |
| 5 | G9078 | [5fjj.1.A](https://swissmodel.expasy.org/templates/5fjj.1) | 0.67 | -3.61 | 38.76% |
| 6 | G9081 | [4iib.1.A](https://swissmodel.expasy.org/templates/4iib.1) | 0.73 | -3.06 | 44.02% |

**Figure S1:** GH3 BGL sequence from *E. nidulans* showing the regions identified in *A. unguis* BGL3 through MALDI TOFF Analysis (red)

1 MAHRWLILAL VAAAAPRALA SPGPSLNERQ SDDEPFSPPY YPAPNGGWVS

51 TWAEAYEKAH SIVSNLTLAE KVNLTTGTGI FMGPCAGQTG SVPRLGIPNL

101 CLHDSPLGVR NTDHNTAFPP GITVGATFDK SLMYERGVGL GEEARGKGVN

151 VLLGPSVGPL GRKPRGGRNW EGFGFDPVLQ GIGGAETIKG MQSTGLIACI

201 KHFVGNEQEM HRMSSVVTQG YSSNIDDRTL HELYIWPFAE GVRAEVGSVM

251 IAYNDVNKSS CSQNSKLING VLKDELGFQG FVVTDWLAHY GGVSSALAGL

301 DMDMPGDGAV PLFGNSYWGP ELSRSILNGT VPVERLNDMV TRILATWYKM

351 GQDQDYPLPN FSSNTEDEKG LLYPGAVISP IGVVNQYVNV QGNHNITARA

401 IARDAITLLK NEGDLLPLRR NDSLKVFGTD AGPDPQGLNS CADKGCNRGV

451 LTMGWGSGTS KLPYLITPQE AIANITPTAE FFITDSFPSS VDANDEDIAI

501 VFINSDSGEN YITVDGNPGD RKTSGLHAWH NGDELVKAAA ERFSQVVVVI

551 HTVGPIILEE WIDLDSVKAV LIAHLPGQEA GYSLTDVLFG DYSPSGHLPY

601 TIPYQESNYP SSVGLLQQPF GQIQDYYTEG LYIDYRHFLK EDITPRYAFG

651 HGLSYTTFEF SEPALSVVTP LDSAYPPSRP AKGPTPTYPN TIPPASEAAW

701 PAKFNRIWRY IYPYLNNPQA DAAVANSSKT YPYPDGYSTD PQPPPRAGGA

751 EGGNPALWDV AFSVQVTVTN TGQHSGRAVA QLYVELPDSL GLDTPSRQLR

801 QFEKTKVLET GQSETLTLEV TRKDVSVWDV EVQDWKTVVG GEGVKIHIGE

851 SVLDIRTECE VGGRCVTL

**Figure S2:** Sequence alignment of peptides P1 and P2 obtained through MALDI-TOF-MS with the translated peptide sequence of gene g2944 in *A. unguis*



**Figure S3:** Alignment of the sequences of PCR amplicons of catalytic domain (cat) of GT BGL, Reverse translated nucleotide sequence of GT-BGL peptide from Mass Spectrometry analyses with full length gene sequence of GT-BGL (g2944)

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 ATGGCTCGCTGGTTATCTATTGCTCTTCTAGCTGCAACGGCACCGTGCAGCCTGGCTCTG

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 CCCGGGGGTAGCCTTGATGCGCGAGCCTCGGACGATGCTTTCTCTCCCCCGTACTATCCA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GCCCCAAAAGGGGGCTGGGTGTCGAGCTGGGCCGATGCGTACCGCAAAGCGCAAGACGTC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GTGGGTAATTTGACACTGGCCGAGAAGGTCAACCTGACGACTGGGACCGGAATTTACATG

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GGCCCCTGTGCGGGCCAAACGGGCAGCGTACCTCGATTCGGCATCCCGAATATTTGTCTT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 CATGATTCTCCCCTCGGAGTGCGCAATGCTGATCACAACACGGCCTTTCCGCCAGGTATC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 ACGGTTGGTGCTACGTTTGACAAGGACCTCATGTACGAGCGCGGGGTAGGACTCGGTGAA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GAAGCACGTGGGAAAGGTGTGAATGTCCTACTTGGTCCCTCTGTCGGCCCTATGGGACGC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 AAACCTAGGGGTGGCCGCAACTGGGAGGGGTTTGGGTTTGATCCTGTTTTGCAGGCCATC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GGTGGTGCGGAGACGATCAAGGGCATGCAGAGTACAGGAGCTATTGCCTGCCTCAAGCAT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TTCATTGGAAATGAGCAGGAGATGCACCGTATGAGCAGTGTCGTCACACAGGGTTACTCG

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TCCAACATTGATGATAGGACTCTGCATGAGTTCTACTTATGGCCTTTCGCCGAAGGTGTC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 CGGGCTGAAGTAGGCTCGGTGATGATTGCTTATAATGATGTAAATATACCTTCCCTGACA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TAGTGGCGTGGCTAATGATTCTAGGTGAACAAATCGGCGTGTAGTCAGAATAGCAAGCTC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 ATCAATGGGGTTCTGAAAGATGAGCTGGGCTTCCAAGGCTTCGTTGTGACTGACTGGCTG

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GCGCACCTCGGTGGAGTTTCGTCTGCCCTGGCGGGTTTGGATATGAGCATGCCTGGAGAT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GGGGCGATTCCATTATTTGGCGACAGTTACTGGGGCCCTGAATTGTCGAGATCTATTCTC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 AACGGATCGGTTCCAGTCGATCGATTGGATGACATGGTATGCCATATGACCAAATGTGCG

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 AACTACAGCTGATTCACCAGGTTACTCGTATTCTCGCAACATGGTTTAAAATGGGCCAGG

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 ATCTGGACTACCCGCTTCCGAACTTTTCTAGCAACACGGACGATGAAACAGGCCCTCTTT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 ATCCAGGCGCTCTTGTGTCCCCCACTGGCGTTGTCAACCAGTACATCAATGTTCAAGGCG

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 ACCACAACATTACTGCCAGAGCTATTGCCAGAGATGCAATCACGCTTCTCAAAAATGAAG

cat -----------------AACCGTAACGACTCATTGAAGGTTTTTGGGACGGATGCTGGAC

P1 ------------------------------------------------------------

g2944 GGGGCGTGCTGCCCCTGAACCGTAACGACTCATTGAAGGTTTTTGGGACGGATGCTGGAC

cat CGGMCCCGAAAGGTTTGAACTCTTGCGGGGACAAGGGTTGCAATAGAGGCGTGCTCACCA

P1 -----------------------------------GGTTGCAATAGAGGCGTGCTCACCA

g2944 CGGACCCGAAAGGTTTGAACTCTTGCGGGGACAAGGGTTGCAATAGAGGCGTGCTCACCA

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

cat TGGGATGGGGAAGCGGTACATCAAAGCTTCCGTACCTGATTACACCACAGGAGGCAATCG

P1 TGGGATGGGGAAGCGGTACATCAAAGCTTCCGTACCTGATTACACCACAGGAGGCAATCG

g2944 TGGGATGGGGAAGCGGTACATCAAAGCTTCCGTACCTGATTACACCACAGGAGGCAATCG

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

cat GCAACATATCCTCTACAGCAGAGTTTTATATCACCGACAGCTTTCCTTCGTCTGTGGAAG

P1 GCAACATATCCTCTACAGCAGAGTTTTATATCACCGACAGCTTTCCTTCGTCTGTGGAAG

g2944 GCAACATATCCTCTACAGCAGAGTTTTATATCACCGACAGCTTTCCTTCGTCTGTGGAAG

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

cat CCGGTGACGACGATATCGCGATTGTCTTTATCAACTCCGACTCTGGCGAGAATTACATTA

P1 CCGGTGACGACGATATCGCGATTGTCTTTATCAACTCCGACTCTGGCGAGAATTACATTA

g2944 CCGGTGACGACGATATCGCGATTGTCTTTATCAACTCCGACTCTGGCGAGAATTACATTA

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

cat CTGTTGATGGCAACCCCGGCGATAGGAAAGCGTCAGGGTTGTACGCCTGGCACGACGGTG

P1 CTGTTGATGGCAACCCCGGCGATAGG----------------------------------

g2944 CTGTTGATGGCAACCCCGGCGATAGGAAAGCGTCAGGGTTGTACGCCTGGCACGACGGTG

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

cat ACGAACTCGTCAAGGCTGCTGCGGAAAAGTTCTCCCAGGTTGTCGTGGTAATTCACACCG

P1 ------------------------------------------------------------

g2944 ACGAACTCGTCAAGGCTGCTGCGGAAAAGTTCTCCCAGGTTGTCGTGGTAATTCACACCG

cat TTGGTCCCATAATTCTCGAGGAGTGGATTGATTTGGATCCGGTCAAGGCTGTCCTTATTG

P1 ------------------------------------------------------------

g2944 TTGGTCCCATAATTCTCGAGGAGTGGATTGATTTGGATCCGGTCAAGGCTGTCCTTATTG

cat CTCACCTACCCGGCCAGGAAGCCGGCTATTCTTTGACTGATATTCTGTTTGGCGCTCATA

P1 ------------------------------------------------------------

g2944 CTCACCTACCCGGCCAGGAAGCCGGCTATTCTTTGACTGATATTCTGTTTGGCGCTCATA

cat GTCCAAGTGGGCATCTCCCCTACACGATTCCCTACAAAGAGTCAGACTACCCCTCGAGTG

P1 ------------------------------------------------------------

g2944 GTCCAAGTGGGCATCTCCCCTACACGATTCCCTACAAAGAGTCAGACTACCCCTCGAGTG

cat TAGGACTGCTCCAACAAGCCTTTGGCCAAATCCAGGACGACTATACAGAGGGTCTCTACA

P1 ------------------------------------------------------------

g2944 TAGGACTGCTCCAACAAGCCTTTGGCCAAATCCAGGACGACTATACAGAGGGTCTCTACA

cat TCGACTACCGCCATTTCTTGAAGGAGGAAATCGCACCACGCTACCCCTTCGGACATGGCC

P1 ------------------------------------------------------------

g2944 TCGACTACCGCCATTTCTTGAAGGAGGAAATCGCACCACGCTACCCCTTCGGACATGGCC

cat TTTCCTACACAACCTTTGAATTCTCAGAGCCCACGTTATCCATTGTCACCCCTCTGGACA

P1 ------------------------------------------------------------

g2944 TTTCCTACACAACCTTTGAATTCTCAGAGCCCACGTTATCCATTGTCACCCCTCTGGACA

cat GCGAATACCCTGCTGCTCGTCCCGACAGGGGCTCTACACCAACATACCCAGATACAATCC

P1 ------------------------------------------------------------

g2944 GCGAATACCCTGCTGCTCGTCCCGACAGGGGCTCTACACCAACATACCCAGATACAATCC

cat CACCCGCATCCGAAGCCGCTTGGCCAAAAAATTTCAACCGTATCTGGCGATATCTATACC

P1 ------------------------------------------------------------

g2944 CACCCGCATCCGAAGCCGCTTGGCCAAAAAATTTCAACCGTATCTGGCGATATCTATACC

cat CCTACCTCGACAACCCCAATGGCGCCACCGCCAACTCTTCAAAGACTTATCCCTACCCTG

P1 ------------------------------------------------------------

g2944 CCTACCTCGACAACCCCAATGGCGCCACCGCCAACTCTTCAAAGACTTATCCCTACCCTG

cat ACGGTTACAGCACAGACCCCAAGCCTCTTCCGCGCGCCGGTGGAGGAGAAGGCGGGAATC

P1 ------------------------------------------------------------

g2944 ACGGTTACAGCACAGACCCCAAGCCTCTTCCGCGCGCCGGTGGAGGAGAAGGCGGGAATC

cat CTGCTCTATGGGATGTTGCATTCTCTGTTCAGGTGACTGTCATGAATTCAGGCTCGCGGT

P1 ------------------------------------------------------------

g2944 CTGCTCTATGGGATGTTGCATTCTCTGTTCAGGTGACTGTCATGAATTCAGGCTCGCGGT

cat CTGGTCGTGCAGTAGCGCAGCTTTACGTTGAACTGCCTGAGTCGGTGGGTCTCGATACGC

P1 ------------------------------------------------------------

g2944 CTGGTCGTGCAGTAGCGCAGCTTTACGTTGAACTGCCTGAGTCGGTGGGTCTCGATACGC

cat CGAGTAGGCAGTTGAGGCAGTT--------------------------------------

P1 ------------------------------------------------------------

g2944 CGAGTAGGCAGTTGAGGCAGTTTGAAAAGACAAAGCTGCTGGAGCCTGGTGACAGTGAGA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 CTGTCACCATGGATATCACGCGGAAGGATGTGAGCGTTTGGGATGTCGTCGTGCAGGACT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GGAGGACAGTTGTTGGCGGAGAGGGCGTTAAGATTTATGTTGGGGAGAGTGTACTTGATA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TGCGGATTGAGTGTGAGGCCGGCGGGGAATGCTCTACAATATAATAGTCTATCATTGTAC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TTATTTTCTTGAAATTATTGTCTGTTCTTCATGATAACTTCCTCTATCATCCTGTCGCGA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GAGCATCACATCCTCCGTATGTATGAGTGCACTAAGTTAAGTATATACATAAGCTGAGTA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 ATCTGCATAACGGCCCACCTAGCTATGTAGTTATCGCACCGACGCTGATCGGCCGACGAT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TTCAACCTTCTGGATCGCAGGTCGACATTAACGTCACGCAGAATTTGAGCGCACAATCTC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 CAAATGCGTGGCACTAGCTCATACTACCGACAACAGCCCTAGCATATTTTCCCTTCAAAG

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 CCTACCATAACGCGAATCTTCACAACAGAAATACCGCCAGGATGCAGGGCTTCAGTACGT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TCAACACCCTTTCCTATCGCAAAGCAAAGCACAAAGAACTAACCATGCTTTCTCGCAGAT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 ATGGGGCGGTACGTTCATACACCCTCTTCCCATTCCCGTATACCTTCCCAATTCGGCACA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 CTCATCGATTTATAGGTACATCCCGCCCGACCTCGAAGGTACAACGACACCCAACAAACT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 CGCCGGCAAGCACCCGCTCGGCGCACGCGCGCGACACCTGCACACAAAGGGAGCACTGAT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 AGTGCGCTTCGAGATGCCCTTTGCGACATGGTGCACGAACTGCGTCCCCGAAGAAATCAT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 CGGGCAGGGCGTGCGCTTCAATGCGGAGAAGAAGAAAGTAGGGAACTACTATTCGACCCC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GATATACAGTTTCCGGTTCAAACATACGGTTTGCGGGAAGTGGATTGAGATCCGCACGGA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TCCGAAGAATACGGACTATGTTGTTACGGAGGGCGGCAGGAGGAGGGATACGGGGGAGGA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TAAGGATGGAGAGAGGGAGGGGGAGATTGTGCTGGGTGGAGTTAGGGCTGGGGGCTCTGC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 AGCCGGGGCTGATGATCCGTTTGCGAAGATTGAAACGAAGGTTGAAGATAAGAGGGCTGT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TGATCAGTCGAAGACGAGGATATTGGAGCTGCAGAAGCGCCAGGCGAGGGACTGGGATGA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TCCATATGAAGTCTCGAGACGGCTGAGGCGCGGGTTCCGGGCTGAGAGGAAGACGTTGGA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GAAGGAGGAGGGGAGGAAAGAGGCGCTGAAGGATAAGATGAGTTTGGGGATTGAGATTGT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GGATGAGCGCGAGGAGGATGGGCTTAGGGCTGCGATGGTTGAGTTTGACGAGGCGGGTGG

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TCGCAGCGGTGGGAATGGGAGTTCCGGGCTGCGTTCGAAGCCGATGTTTGAGACCAAGGC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TCTTATAGGAACCAAATCTGGTGATAAGAAGAAGAATGGGAAGCGAAAACCGGCTGACCT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 CATTGCCGAGCGGAAGGCGATGTTTCGCAGTGAGTTGGCTGGGAATACGCGTGCCGCTGT

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TGATCCATTTTTGAATGGTGATACTACAAGTGCCTGGGAGCCGGAGGTCAAAAAGAGGAA

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 GGGCTCGAATAAGGCCTGCTCAGAGGGTAAGAAAGACGGTGAGCCCGTAAACGGTGATGC

cat ------------------------------------------------------------

P1 ------------------------------------------------------------

g2944 TTCTCAATCTACAAGTAATTGCCCAGGACCGCCGAAAGGTGCAAGCGAAGAATCTAAGCA

cat -------------------------------------------------

P1 -------------------------------------------------

g2944 GCAGAACACCTCTGTTTCCCTGGTGGACTACGGATCTGACTCCGAGTGA